Example of Frisch-Waugh Theorem

The Frisch-Waugh theorem says that the multiple regression coefficient of any single variable can also be obtained by first netting out the effect of other variable(s) in the regression model from both the dependent variable and the independent variable.

$$\hat{\beta}_1 = (X_1'M_2'M_2X_1)^{-1}X_1'M_2'M_2y = (X_1'X_1)^{-1}X_1'y - (X_1'X_1)^{-1}X_1'X_2\beta_2$$

In practice this means regressing the residuals when y is regressed on X_2 , (M_2y), on the residuals from when X_1 is regressed on X_2, (M_2X_1)

First regress the dependent variable on the X_2 variable (in this case London)

```
. reg grosspay london

      Source |       SS       df       MS              Number of obs =   12266
-------------+------------------------------           F(  1, 12264) =   18.73
       Model |  2998066.02     1  2998066.02           Prob > F      =  0.0000
    Residual |  1.9629e+09 12264  160055.968           R-squared     =  0.0015
-------------+------------------------------           Adj R-squared =  0.0014
       Total |  1.9659e+09 12265  160287.359           Root MSE      =  400.07

------------------------------------------------------------------------------
    grosspay |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      london |   50.70394    11.7154     4.33   0.000     27.73992    73.66797
       _cons |   232.5472   3.821297    60.86   0.000     225.0568    240.0375
------------------------------------------------------------------------------

. predict uhat1, resid
```

and save the residuals (named uhat1). This gives M_2y

Next regress the X_1 variable on the X_2 variable and save these residuals (named uhat2). This gives M_2X_1

```
. reg age london

      Source |       SS       df       MS              Number of obs =   12266
-------------+------------------------------           F(  1, 12264) =   21.36
       Model |  2681.55785     1  2681.55785           Prob > F      =  0.0000
    Residual |  1539528.39 12264  125.532322           R-squared     =  0.0017
-------------+------------------------------           Adj R-squared =  0.0017
       Total |  1542209.95 12265  125.740722           Root MSE      =  11.204

------------------------------------------------------------------------------
         age |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      london |  -1.516403   .3280945    -4.62   0.000     -2.15952   -.8732865
       _cons |   40.01296    .107017   373.89   0.000     39.80318    40.22273
------------------------------------------------------------------------------

. predict uhat2, resid
```

Now regress the residuals ${}_2y$ on the residuals M_2X_1

```
. reg uhat1 uhat2

      Source |       SS       df       MS              Number of obs =   12266
-------------+------------------------------           F(  1, 12264) =    2.96
       Model |  473241.472     1  473241.472           Prob > F      =  0.0855
    Residual |  1.9625e+09 12264  160017.382           R-squared     =  0.0002
-------------+------------------------------           Adj R-squared =  0.0002
       Total |  1.9629e+09 12265   160042.92           Root MSE      =  400.02

------------------------------------------------------------------------------
       uhat1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       uhat2 |   .5544311   .3223961     1.72   0.086     -.077516    1.186378
       _cons |   2.02e-07    3.61187     0.00   1.000    -7.079833    7.079834
```

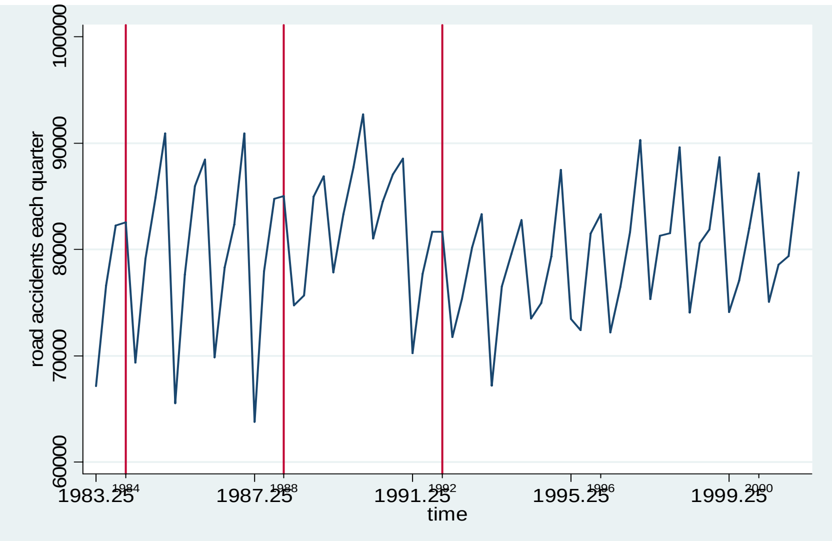
and check that this is the same as in the multiple regression (it is)

```
. reg grosspay age london

      Source |       SS       df       MS              Number of obs =   12266
-------------+------------------------------           F(  2, 12263) =   10.85
       Model |  3471307.55     2  1735653.78           Prob > F      =  0.0000
    Residual |  1.9625e+09 12263  160030.429           R-squared     =  0.0018
-------------+------------------------------           Adj R-squared =  0.0016
       Total |  1.9659e+09 12265  160287.359           Root MSE      =  400.04

------------------------------------------------------------------------------
    grosspay |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .5544311   .3224092     1.72   0.086    -.0775417    1.186404
      london |   51.54468   11.72466     4.40   0.000      28.5625    74.52687
       _cons |   210.3628   13.45452    15.64   0.000     183.9898    236.7357
```

In practice you will hardly ever use the residual approach to obtain multiple regression coefficients (one exception to this is if the X_2 variables capture seasonal or time effects, in which case you can think of the 2-step process as an alternative method of “de-seasonalising” or “detrending” both dependent and control variables such as exhibit strong trends or seasonality as in the road accidents data below). The usefulness of this exercise is that it helps makes clearer the partialing out nature of multiple regression.



v

twoway (line acc time), xlabel(1983.25(4)2000) xmlabel(1984(4)2000) xline(1984 1988 1992) ytitle(road accidents each quarter)

Another way to think abour partialling out is to recognise that the multivariate OLS estimate of any single variable can always be written as

$$\hat{\beta}_k = \frac{Cov(M_2 X_k, Y)}{Var(M_2 X_k)}$$

(see lecture notes and exercise 1)

Consider an OLS regression of food share expenditure on total expenditure and age

```
use "C:\qea\food.dta", clear
```

The multivariate regression model is given by

```
. reg  foodsh expnethsum age

      Source |       SS       df       MS              Number of obs =     200
-------------+------------------------------           F(  2,   197) =   29.37
       Model |  4290.50386     2  2145.25193           Prob > F      =  0.0000
    Residual |  14387.5993   197  73.0334989           R-squared     =  0.2297
-------------+------------------------------           Adj R-squared =  0.2219
       Total |  18678.1031   199  93.8598148           Root MSE      =   8.546

------------------------------------------------------------------------------
      foodsh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  expnethsum |  -.0128509   .0021607    -5.95   0.000    -.0171119   -.0085898
         age |   .1023938   .0418339     2.45   0.015      .019894    .1848935
       _cons |   18.78057   2.657014     7.07   0.000     13.54072    24.02041
------------------------------------------------------------------------------
```

We are interested in understanding the estimated OLS coefficient on age in a multiple regression

Partialling out the effect of total expenditure on the food share

```
. reg  foodsh expnethsum

      Source |       SS       df       MS              Number of obs =     200
-------------+------------------------------           F(  1,   198) =   51.46
       Model |  3852.96953     1  3852.96953           Prob > F      =  0.0000
    Residual |  14825.1336   198  74.8744122           R-squared     =  0.2063
-------------+------------------------------           Adj R-squared =  0.2023
       Total |  18678.1031   199  93.8598148           Root MSE      =   8.653

------------------------------------------------------------------------------
      foodsh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  expnethsum |  -.0147015   .0020494    -7.17   0.000     -.018743     -.01066
       _cons |   24.87616   .9376504    26.53   0.000      23.0271    26.72522
------------------------------------------------------------------------------
```

saving the residual

```
. predict yhat2 if e(sample), resid
```

Partialling out the effect of total expenditure on age

```
. reg  age expnethsum

      Source |       SS       df       MS              Number of obs =     200
-------------+------------------------------           F(  1,   198) =   27.63
       Model |  5823.33649     1  5823.33649           Prob > F      =  0.0000
    Residual |  41731.6185   198   210.76575           R-squared     =  0.1225
-------------+------------------------------           Adj R-squared =  0.1180
       Total |   47554.955   199  238.969623           Root MSE      =  14.518

------------------------------------------------------------------------------
         age |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  expnethsum |  -.0180738   .0034385    -5.26   0.000    -.0248545   -.0112931
       _cons |   59.53094   1.573165    37.84   0.000     56.42863    62.63324
------------------------------------------------------------------------------
```

and saving this residual

```
. predict xhat2 if e(sample), resid
```

then regressing the first residual on the second (in a univariate model) we see the estimated OLS estimate on age from the multiple regression above

```
. reg  yhat2 xhat2

      Source |       SS       df       MS              Number of obs =     200
-------------+------------------------------           F(  1,   198) =    6.02
       Model |  437.534334     1  437.534334           Prob > F      =  0.0150
    Residual |  14387.5994   198  72.6646436           R-squared     =  0.0295
-------------+------------------------------           Adj R-squared =  0.0246
       Total |  14825.1338   199  74.4981597           Root MSE      =  8.5244

------------------------------------------------------------------------------
       yhat2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       xhat2 |   .1023938   .0417281     2.45   0.015     .0201051    .1846824
       _cons |   6.26e-09    .602763     0.00   1.000    -1.188659    1.188659
------------------------------------------------------------------------------

. predict ahat if e(sample)
(option xb assumed; fitted values)

. sort xhat2
```

Graphing this relationship

```
. two (scatter  yhat2 xhat2) (line ahat xhat2)
```